IS 5002 (1969): Method for determination of sample size to estimate the average quality of a lot or process [TXD 1: Physical Methods of Tests]

BLANK PAGE

Indian Standard

METHODS FOR DETERMINATION OF SAMPLE SIZE TO ESTIMATE THE AVERAGE QUALITY OF A LOT OR PROCESS

(Fifth Reprint MARCH 1993)

UDC 620.115.8

BUREAU OF INDIAN STANDARDS
MANAK BHAVAN, 9 BAHADUR SHAH ZAFAR MARG
NEW DELHI 110002

April 1970

Indian Standard
METHODS FOR DETERMINATION OF SAMPLE SIZE TO ESTIMATE THE AVERAGE QUALITY OF A LOT OR PROCESS

Sampling Methods Sectional Committee, TDC 33

Chairman	*Representing*
Dr (Kumari) S. P. Vaswani	N. C. Corporation Private Ltd, Bombay
Members	
Dr Daroga Singh	Institute of Agricultural Research Statistics (ICAR), New Delhi
Shri V. N. Amble (*Alternate*)	
Dr Jai Prakash	Cotton Technological Research Laboratory (ICAR), Bombay
Shri G. S. Rajaraman (*Alternate*)	
Shri P. Krishnaswami	Ministry of Defence (R & D), New Delhi
Dr N. T. Mathew	Indian Statistical Institute, Calcutta
Shri B. K. Raja Rao (*Alternate*)	
Dr C. Nanjundayya	The Bombay Textile Research Association, Bombay
Shri M. V. Natarajan	Education Department, Government of Mysore, Bangalore
Shri I. L. Patel	Inspection Wing, Directorate General of Supplies & Disposals, New Delhi
Shri A. K. Pramanick (*Alternate*)	
Shri R. S. Prayag	Textiles Committee, Bombay
Shri B. Raha	The Bombay Dyeing and Mfg Co Ltd, Bombay
Shri S. S. Iyer (*Alternate*)	
Shri T. V. Ratnam	The South India Textile Research Association, Coimbatore
Shri K. N. Seshan (*Alternate*)	
Representative	Aryodaya Spg & Wvg Co Ltd, Ahmedabad
Representative	The Ahmedabad Textile Industry's Research Association, Ahmedabad
Shri N. V. Ullal	Khatau Makanji Spg & Wvg Co Ltd, Bombay
Shri S. D. Mathure (*Alternate*)	
Dr B. N. Singh, Director (Stat)	Director General, ISI (*Ex-officio Member*)

Secretary

Shri G. W. Datey
Assistant Director (Stat), ISI

BUREAU OF INDIAN STANDARDS
MANAK BHAVAN, 9 BAHADUR SHAH ZAFAR MARG
NEW DELHI 110002

Indian Standard

METHODS FOR DETERMINATION OF SAMPLE SIZE TO ESTIMATE THE AVERAGE QUALITY OF A LOT OR PROCESS

0. FOREWORD

0.1 This Indian Standard was adopted by the Indian Standards Institution on 17 October 1969, after the draft finalized by the Sampling Methods Sectional Committee had been approved by the Textile Division Council.

0.2 The estimation of the average quality of a lot of material or a process is of great importance to the consumer and the producer. The average quality can be found by inspecting all the items in the lot or alternatively it can be estimated by inspecting a part of the lot selected in a manner so as to constitute a representative sample. Hundred percent inspection, unlike the sampling inspection, is generally uneconomical, in some cases impracticable and in many cases impossible, as in case of destructive tests. In case of sampling inspection, however, an important problem is to determine the size of the sample which will enable the estimation of the average quality with a specified degree of accuracy. The methods given in this standard are expected to be helpful in this respect.

0.3 It may be clarified that the sample sizes specified in sampling clauses included in various Indian Standards are designed for the purpose of accepting or rejecting a lot, while the sample sizes as obtained by the use of the methods given in this standard are meant for estimation of the average quality of the lot. As the samples are drawn with different purposes in the two cases, the sample sizes are likely to be different.

0.4 This standard is one of a series of Indian Standards relating to the techniques of the statistical quality control. Other standards published so far in the series are:

- IS : 397-1952 Method for statistical quality control during production by the use of control chart (*tentative*)
- IS : 1548-1960 Manual on basic principles of lot sampling
- IS : 2500 (Part I)-1963 Sampling inspection tables : Part I Inspection by attributes and by count of defects
- IS : 2500 (Part II)-1965 Sampling inspection tables : Part II Inspection by variables for percent defective
- IS : 4905-1968 Methods for random sampling

0.5 In preparing this standard, considerable assistance has been derived from ASTM Designation : E 122-58 ' Recommended practice for choice of

sample size to estimate the average quality of a lot or process' issued by the American Society for Testing and Materials.

0.6 In reporting the result of a test or analysis, if the final value, observed or calculated, is to be rounded off, it shall be done in accordance with IS : 2-1960*. However, the value of the sample size, obtained by the use of any of the formula given in this standard, shall be rounded up so as not to reduce the accuracy of the estimate.

---

1. SCOPE

1.1 This standard lays down methods for calculating the size of the sample required to estimate, with a specified limit of error and probability level, the average quality of a lot or process.

2. TERMINOLOGY

2.0 For the purpose of this standard, the following definitions shall apply.

2.1 Attribute — A qualitative characteristic according to which an item is classified as defective or non-defective.

2.2 Average Quality — In the case of attributes, it refers to the percentage or proportion of defectives; in the case of variables, it refers to the mean value of the individual measurements on items measured.

2.3 Coefficient of Variation — The standard deviation expressed as percentage of the mean value of the test results.

2.4 Defective — An item, the quality of which does not meet the specified requirements for the characteristic(s) under consideration.

2.5 Limit of Error of Estimate — The maximum difference between the estimate (to be made on the basis of a sample) and its true value (that would be obtained if all the items in the lot or process were tested) at a given probability level.

2.6 Lot — A batch, a group, a continuous stream or a bulk of product or raw material submitted for inspection.

2.7 Probability Level — A measure of probability associated with the sample size and limit of error of estimate. It expresses the probability that the difference between the estimate based on the sample of a particular size and its true value does not exceed the specified limit of error.

2.8 Sampling in Stages — The method of selecting a sample in which the sampling units at each stage are sub-sampled from the (larger) units chosen at previous stage. The sampling units pertaining to the first stage are called first-stage units; similarly second-stage units, and so on.

*Rules for rounding off numerical values (*revised*).

2.9 Standard Deviation — The square-root of the quotient obtained by dividing the sum of squares of the deviations (differences) of the individual test results from their mean value, by the number of test results (*see* Appendix A for the method of calculation of standard deviation).

2.10 Variable — A quantitative characteristic according to which the value of the characteristic for any item is measured on a continuous scale and is expressed in terms of the units of measurement.

3. DETERMINATION OF SAMPLE SIZE WHEN EMPIRICAL KNOWLEDGE IS AVAILABLE

3.1 Empirical Knowledge Available

3.1.1 For correct determination of the sample size, it is extremely useful if some knowledge of the variability in the lot or process is available. Such information can usually be obtained from the analysis of past data or from a study of process capability. The empirical knowledge that is useful can be of the following types:

a) In case of attributes, since the distribution is completely defined when the percent defective is known, an estimate of the percent defective gives all the information that is needed.

b) In case of variables, knowledge of the standard deviation or coefficient of variation is required. In cases where two-stage sampling is required, knowledge of the standard deviation between and within the first-stage units would be needed.

c) If the standard deviation is not known, the range or spread of the characteristic from its lowest to the highest value, together with the shape of the distribution of the values, would be desirable.

3.1.2 Even if the empirical knowledge is meagre, it may be used to determine the sample size. On the basis of the test results obtained from a sample of such a size, the adequacy of the sample size may be ascertained. If the size of the sample already drawn is found to be not adequate, additional items may be drawn, if convenient, and tested so as to get an estimate with the desired limit of error. In any case, the test results already obtained will furnish an estimate of the variability in the lot which may be used for ascertaining the size of the sample to be drawn in future.

3.2 Sampling for Attributes

3.2.1 When the characteristic under consideration is of attribute type, the error limit can be specified as the maximum difference in the true value of percent defective and its estimate that can be tolerated.

3.2.2 For estimating the percent defective in the lot or process, some prior knowledge of the percent defective is desirable for determination of

the sample size. The sample size (n) in such cases is given by the formula:

$$n = \frac{u^2 p(100 - p)}{d^2}$$

where

u = factor corresponding to the selected probability level (*see* Table 1),

p = estimate of the percent defective from past data, and

d = limit of error expressed as the difference in the true percent defective and its estimate.

TABLE 1 VALUES OF *u*

(*Clause* 3.2.2)

Probability Level percent	u
99	2·58
98	2·33
95	1·96
90	1·64

Example 1:

It is known from previous experience that a manufacturing process produces about 10 percent defective pirns in respect of concentricity requirement. It is required to determine the minimum number of pirns that should be subjected to the concentricity test if it is desired to obtain an estimate of the percent defective with an error limit of ±2 percent and the probability level of 95 percent.

From Table 1 we have u = 1·96, and from previous experience we know that p = 10 percent. Since d = 2 percent, n is obtained as:

$$n = \frac{(1{\cdot}96)^2 \times 10 \times (100 - 10)}{2^2}$$

$$= \frac{3{\cdot}841\,6 \times 900}{4}$$

$$= 864{\cdot}36 \text{ or } 865.$$

It is, therefore, necessary to examine at least 865 pirns in order to obtain an estimate of the percent defective with the error limit of ±2 percent at 95 percent probability level.

3.2.3 *Estimation of Percent Defective from Previous Data* — If data on a number of lots from the same source or process are available, such information could be used for obtaining an estimate of the percent defective. For instance, if k lots of size $n_1, n_2, \ldots n_k$ have been inspected previously and $m_1, m_2 \ldots m_k$ defectives were found as a result of inspection, then the estimate of the percent defective (p) is given by:

$$p = \frac{m_1 + m_2 + \ldots\ldots\ldots\ldots + m_k}{n_1 + n_2 + \ldots\ldots\ldots\ldots + n_k} \times 100.$$

The value so obtained may be used in the formula given in **3.2.2** to estimate the sample size.

Example 2:

It is required to ascertain the sample size needed to estimate the percent defective in a lot of glass bottles to be inspected with respect to visual defects like cords, blisters, bubbles, stones when the limit of error of estimate is ± 3 percent with the probability level of 95 percent. The number of glass bottles inspected and the number of defectives found in the four lots tested previously are as given below:

Lot No.	1	2	3	4
Number Inspected	75	100	90	125
Number of Defectives	3	10	6	11

From the available data, the estimate of the percent defective (p) is obtained as:

$$p = \frac{3 + 10 + 6 + 11}{75 + 100 + 90 + 125} \times 100$$

$$= \frac{3000}{390}$$

$$= 7{\cdot}7$$

From Table 1 we have $u = 1{\cdot}96$; and since $d = 3$, the sample size is obtained as:

$$n = \frac{(1{\cdot}96)^2 \times 7{\cdot}7 \times (100 - 7{\cdot}7)}{(3)^2}$$

$$= \frac{3{\cdot}841\,6 \times 7{\cdot}7 \times 92{\cdot}3}{9}$$

$$= 303{\cdot}3 \text{ or } 304$$

Hence, at least 304 glass bottles will have to be examined for visual defects in order to obtain an estimate of the percent defective with the limit of error of ± 3 percent at the probability level of 95 percent.

3.3 Sampling for Variables

3.3.1 When the characteristic is of a variable type, the error limit can be specified either:

a) in terms of the units in which the characteristic is being measured, or

b) in terms of the percentage of the true mean value of the characteristic.

NOTE — The choice between the two ways (a) and (b) depends on whether the standard deviation or coefficient of variation of the characteristic is constant over the range of possible values of the average. If the standard deviation is constant, the method (a) should be chosen, whereas if the coefficient of variation is constant, method (b) should be chosen.

3.3.2 When the standard deviation is known, the sample size (n) is given by the formula:

$$n = (u\, s/E)^2$$

where

u = factor corresponding to the selected probability level (*see* Table 1),

s = known standard deviation, and

E = error limit expressed in the same units of measurement as the standard deviation.

Example 3:

It is required to find the sample size needed to estimate the resistivity of rolled aluminium rods for electrical purposes when the limit of error of the estimate is 0·03 micro-ohm centimetre (μΩ-cm) and the standard deviation is known to be 0·07 μΩ-cm with the probability level of 95 percent.

Since the probability level is 95 percent, we have from Table 1, u = 1·96; and since s = 0·07 μΩ-cm and E = 0·03 μΩ-cm, we get by substituting in the formula:

$$n = \left(\frac{1{\cdot}96 \times 0{\cdot}07}{0{\cdot}03} \right)^2$$
$$= (4{\cdot}57)^2$$
$$= 20{\cdot}88 \text{ or } 21$$

Hence, the minimum number of tests required to obtain an estimate of the resistivity of the aluminium rods with an error limit of 0·03 μΩ-cm at 95 percent probability level is 21.

3.3.3 In case of some materials where the standard deviation varies approximately as the mean value, it is found that the coefficient of

variation remains constant over a fairly long range of values. In such cases, it is convenient to use the following formula:

$$n = (u\, v/e)^2$$

where

u = factor corresponding to the selected probability level (*see* Table 1),

v = coefficient of variation in percent, and

e = limit of error expressed in percent.

Example 4:

It is required to find the minimum number of tests for ultimate tensile strength of a certain type of wire when the coefficient of variation is known to be 9 percent and the estimate of the strength is required to be within a limit of error of ± 5 percent at 99 percent probability level.

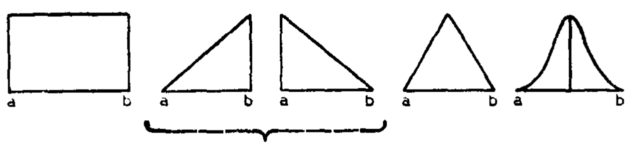
Since the probability level is 99 percent, we have from Table 1, $u = 2{\cdot}58$; and since $v = 9$ percent and $e = 5$ percent, we get by substituting in the formula:

$$n = \left(\frac{2{\cdot}58 \times 9}{5} \right)^2$$
$$= (4{\cdot}64)^2$$
$$= 21{\cdot}52 \text{ or } 22$$

Hence, 22 tests have to be conducted in order to estimate the tensile strength with the limit of error of $\pm$ 5 percent at 99 percent probability level.

3.3.4 *Estimation of Standard Deviation from the Range and Shape of the Distribution* — If no information about the value of standard deviation or coefficient of variation is available, it is not possible to use the formulae given in **3.3.2** and **3.3.3**. However, an approximate estimate of the standard deviation may be obtained from the knowledge of the range of values of the characteristic and shape of distribution. Thus, if most of the values run uniformly from one end to the other (rectangular distribution) or lie at one end (triangular distribution) or mostly in the middle (triangular or normal distribution) the formulae for estimating standard deviation in such situations are given below:

Distribution



Estimate of Standard Deviation

$$\frac{b-a}{3{\cdot}5} \qquad \frac{b-a}{4{\cdot}2} \qquad \frac{b-a}{4{\cdot}9} \qquad \frac{b-a}{6{\cdot}0}$$

The knowledge of the shape of the distribution could normally be obtained from the information concerning the past behaviour of lots or the process, or from the study of the usual procedure of blending, mixing, stacking, storing, etc. In cases of doubt about the shape of the distribution, or if the characteristic under consideration is critical, the estimate of the standard deviation corresponding to rectangular distribution may be used to calculate the sample size. The sample size so obtained will be larger than in other cases, thereby keeping the error of the estimate within the desired limit.

Example 5:

It is required to determine the minimum sample size needed to estimate the width of a cotton fabric with a limit of error of ± 1 cm when it is known from previous experience that the width varies from 70 cm to 80 cm and that the values are clustered around the mid-point of this range but do not follow the normal distribution.

The information available from previous experience suggests that the distribution of values would be approximately like an isosceles triangle. The estimate of the standard deviation (s) is obtained as:

$$s = \frac{10}{4{\cdot}9}$$

$$= 2{\cdot}04 \text{ cm}$$

Assuming the probability level to be 95 percent, we get from Table 1, $u = 1{\cdot}96$. Substituting in the formula given in **3.3.2**, we get:

$$n = \left(\frac{1{\cdot}96 \times 2{\cdot}04}{1} \right)^2$$

$$= (3{\cdot}998)^2$$

$$= 15{\cdot}98 \text{ or } 16$$

Hence, the minimum number of tests necessary to estimate the width of the cotton fabric within $\pm$ 1 cm is 16.

3.3.5 *Estimation of Standard Deviation from Previous Data* — From the previous data on the measurement of the characteristic, it is sometimes possible to obtain an estimate of the standard deviation. The observed values may be arranged in random order and the ranges for successive groups of suitable size of about 4 to 10 values may be calculated. If the average of such ranges is denoted by $\bar{R}$, then $D \times \bar{R}$ is an estimate of the standard deviation, where the values of D for different group sizes are as given in Table 2. The standard deviation so estimated may then be used in the formula given in **3.3.2**.

TABLE 2 VALUES OF *D*

(*Clause* 3.3.5)

GROUP SIZE	*D*
4	0·485 7
5	0·429 9
6	0·394 6
7	0·369 8
8	0·351 2
9	0·336 7
10	0·324 9

Example 6:

It is required to compute the minimum number of tests required to estimate BHC content in BHC dusting powder with 10 percent nominal content with the limit of error of ± 0·2 percent at 95 percent probability level, when the data on 10 batches of the same material is available and is as follows:

Sample No. / *Batch No.*	1	2	3	4	5	*Range*
	BHC Content in Percent					
1	10·09	10·39	10·09	9·78	10·19	0·61
2	10·19	9·68	10·39	10·39	9·68	0·71
3	9·68	9·68	9·98	10·09	9·78	0·41
4	8·38	8·68	8·88	8·98	8·48	0·60
5	9·88	9·88	10·39	10·39	9·88	0·51
6	10·29	10·09	9·88	10·19	10·19	0·41
7	9·78	10·29	10·19	10·29	9·98	0·51
8	10·09	10·09	10·29	9·88	9·78	0·51
9	10·09	9·88	10·19	9·88	10·09	0·31
10	9·88	9·78	9·78	9·88	9·98	0·20
					Total	**4·78**

The range for each batch is shown in the last column of the table given on previous page. The average of all such ranges is 0·478. The factor D corresponding to the group of size 5 as obtained from Table 2 is 0·429 9. The estimate of the standard deviation (s) is obtained as:

$$s = 0{\cdot}429\,9 \times 0{\cdot}478$$
$$= 0{\cdot}205$$

For 95 percent probability level, we get from Table 1, $u = 1{\cdot}96$; and by substituting in the formula given in **3.3.2**, we get:

$$n = \left(\frac{1{\cdot}96 \times 0{\cdot}205}{0{\cdot}2} \right)^2$$
$$= (2{\cdot}009)^2$$
$$= 4{\cdot}04 \text{ or } 5$$

Hence, the minimum number of tests necessary to estimate BHC content in the BHC dusting powder with the limit of error of $\pm$ 0·2 is 5.

3.4 Two-Stage Sampling for Attributes — The method for determination of sample size is the same as that given in **3.2.2**. The number of first-stage units to be selected is, however, determined on considerations of convenience and the expenses involved. When the number of first-stage units is determined on the above considerations, the sample of the required size is obtained by selecting equal number of items for each first-stage unit.

3.5 Two-Stage Sampling for Variables — When the material is supplied in a form where the items cannot be sampled directly, two-stage sampling is generally employed. In such cases, the first-stage unit is some collection of items, for instance, in case of cloth the first-stage units will be the bales of cloth in which a certain number of pieces of cloth are packed. The sampling will then consist of selecting a number of bales from the lot and then from each of the bales so chosen a certain number of pieces of cloth will be selected. All these pieces of cloth will form the sample for the purpose of inspection. For estimating the number of first-stage units and the number of items from each first-stage unit to be drawn in the sample, it is necessary to have some knowledge about the standard deviation between the first-stage units and the standard deviation between the items within the same first-stage unit. If these are known, the sample size (n) of the first-stage units is given by:

$$n = \frac{N (s_w^2 + k s_b^2)}{kN (E/u)^2 + k s_b^2}$$

where

N = total number of first-stage units in the lot,

s_w = standard deviation for the items within the same first-stage unit,

k = number of items to be selected from each first-stage unit,

s_b = standard deviation for the first-stage units in the lot,

E = limit of error expressed in the same units of measurement as the standard deviation, and

u = factor corresponding to the selected probability level (*see* Table 1).

NOTE 1 — In case the values of s_w and s_b are not known, their estimates are obtained by collecting the necessary data on the basis of statistically designed experiment.

NOTE 2 — If the number of first-stage units in the lot is large, the above formula is reduced to:

$$n = \frac{s_w^2 + k s_b^2}{k (E/u)^2}$$

NOTE 3 — In the formulae given above for n, it is necessary to know the value of k. This value may be determined either from practical considerations (*see also* **5.3**) or by the formula given in **3.3.2** taking, if necessary, a larger limit of error (E) pertaining to the standard deviation (s_w) within the same first-stage unit.

NOTE 4 — If $k = 1$, the same formula as in **3.3.2** holds good for two-stage sampling as if the first-stage units were to be drawn at a single stage, and one item is drawn from each first-stage unit included in the sample for the purpose of inspection.

Example 7:

In sampling of wool from packages to estimate the hard scoured wool content, the packages are considered as first-stage units and the portion of the wool obtained with the help of the sampling implement is considered as item. It is desired to find out the number of first-stage units and items from each first-stage unit, as defined above, necessary to be selected so as to obtain an estimate of the hard scoured wool content within ±1·0 percent at the probability level of 95 percent. The number of packages in the lot is 500 and the standard deviation between items within the same first-stage unit is known to be 2·0 and that between first-stage units is 4·0.

If only one item is to be drawn from each first-stage unit, the number of first-stage units to be selected will be given by the formula in **3.5** as:

$$n = \frac{500 [(2{\cdot}0)^2 + 1 (4{\cdot}0)^2]}{1 \times 500 \left(\frac{1{\cdot}0}{1{\cdot}96} \right)^2 + 1 (4{\cdot}0)^2}$$

$$= \frac{500 (4{\cdot}00 + 16{\cdot}00)}{130{\cdot}15 + 16{\cdot}00}$$

$$= \frac{10\,000}{146{\cdot}15}$$

$$= 68{\cdot}4 \text{ or } 69$$

If two items are to be selected, then

$$n = \frac{500\,[\,(\,2{\cdot}0\,)^2 + 2\,(\,4{\cdot}0\,)^2\,]}{2 \times 500 \left(\frac{1{\cdot}0}{1{\cdot}96}\right)^2 + 2\,(\,4{\cdot}0\,)^2}$$

$$= \frac{18\,000}{260{\cdot}3 + 32{\cdot}0}$$

$$= \frac{18\,000}{292{\cdot}3}$$

$$= 61{\cdot}6 \text{ or } 62$$

Since, the number of first-stage units required is not reduced to a great extent, we may choose the combination $n = 69$ and $k = 1$.

4. DETERMINATION OF SAMPLE SIZE WHEN EMPIRICAL KNOWLEDGE IS NOT AVAILABLE

4.1 When no prior information about the variability is available, one may test a few items for the characteristic under consideration and obtain an estimate of the variability. This estimate, though based on a small number of test results, may be used to find out the minimum sample size that would be needed to estimate the average quality with the given limit of error. Formulae for such determinations are given for one-stage sampling only, as estimates for two-stage sampling in such cases are somewhat more complicated. If two-stage sampling is necessary, the sample size obtained on the basis of one-stage sampling may be taken to be the required number of items. These items may then be selected from a convenient number of first-stage units.

4.2 Sampling for Attributes — An initial sample of fairly large size may be selected from the lot and inspected. If n_1 is the initial sample size and m_1 is the number of defectives in the initial sample, then the estimate of the percent defective (p_1) is given by:

$$p_1 = \frac{100 \times m_1}{n_1}$$

The sample size necessary to estimate the percent defective with the given error limit is then given by:

$$n = \frac{u^2\, p_1\,(\,100 - p_1\,)}{d^2}$$

where

u = factor corresponding to the selected probability level (*see* Table 1), and

d = limit of error expressed as the difference in the true percent defective and its estimate.

If the estimate of the sample size (n) is greater than n_1, then $n - n_1$ additional items may be selected and inspected.

4.3 Sampling for Variables — From the lot or process, an initial sample of a small size, say n_1, may be selected and tested for the characteristic. From the test results so obtained, an estimate of the variability in the lot may be ascertained as follows:

$$s_1^2 = \frac{1}{n_1 - 1}\left[(x_1^2 + x_2^2 + \ldots\ldots x_{n_1}^2) - \frac{(x_1 + x_2 + \ldots\ldots x_{n_1})^2}{n_1}\right]$$

where x_1, x_2, ... x_{n_1} denote the test results.

The sample size necessary to obtain an estimate of the average quality with the given limit of error is then given by:

$$n = \frac{t^2 s_1^2}{E^2}$$

where

t = factor corresponding to the initial sample size (n_1) and the probability level (*see* Table 3),

s_1 = estimate of the standard deviation obtained from the initial sample, and

E = limit of error expressed in the same units as the standard deviation.

If the estimate of the sample size (n) so obtained is greater than the initial sample size (n_1), then $n - n_1$ additional items may be selected from the lot or process and tested.

5. CONSIDERATION OF COST

5.1 In the determination of the sample size two considerations are involved, namely, the limit of error of the estimate to be obtained from the sample and the cost of testing the sample. The formulae for the sample size given in **3** and **4** are applicable to the cases, where the limit of error has been specified. However, in certain cases, due to the limitations of the expenditure that can be permitted on testing, it may not be always possible to test a sample of the size given by these formulae. In such cases, the sample size may be determined on the basis of the permissible cost and then the limit of error associated with the sample size may be estimated.

TABLE 3 VALUES OF *t*

(*Clause* 4.3)

Probability Level Percent / Sample Size	90	95	98	99
2	6·314	12·706	31·821	63·657
3	2·920	4·303	6·965	9·925
4	2·353	3·182	4·541	5·841
5	2·132	2·776	3·747	4·604
6	2·015	2·571	3·365	4·032
7	1·943	2·447	3·143	3·707
8	1·895	2·365	2·998	3·499
9	1·860	2·306	2·896	3·355
10	1·833	2·262	2·821	3·250
11	1·812	2·228	2·764	3·169
12	1·796	2·201	2·718	3·106
13	1·782	2·179	2·681	3·055
14	1·771	2·160	2·650	3·012
15	1·761	2·145	2·624	2·977
16	1·753	2·131	2·602	2·947
17	1·746	2·120	2·583	2·921
18	1·740	2·110	2·567	2·898
19	1·734	2·101	2·552	2·878
20	1·729	2·093	2·539	2·861
21	1·725	2·086	2·582	2·845
22	1·721	2·080	2·518	2·831
23	1·717	2·074	2·508	2·819
24	1·714	2·069	2·500	2·807
25	1·711	2·064	2·492	2·797
26	1·708	2·060	2·485	2·787
27	1·706	2·056	2·479	2·779
28	1·703	2·052	2·473	2·771
29	1·701	2·048	2·467	2·763
30	1·699	2·045	2·462	2·756
Above 30	Use the same values as for *u* (Table 1)			

5.2 One-Stage Sampling — In one-stage sampling, the total cost is generally made up of two parts, namely, the overhead administrative cost and the actual cost of sampling and testing the items selected in the sample. The overhead cost is generally constant whereas the cost of testing is directly proportional to the size of the sample. The total cost of inspection

may, therefore, be represented by the formula:

$$\text{Total cost} = C_o + C_1 n$$

where

C_o = overhead cost,

C_1 = cost of sampling and testing a single item, and

n = sample size.

If it is desired that the total costs should not exceed a certain limit, the sample size will be automatically determined. From the sample size so determined and from the knowledge of variability in the material, it will be possible to determine the limit of error of the estimate. On the other hand, if the limit of error of the estimate is fixed, the sample size will be determined by the formulae given in **3**, and then from the knowledge of the overhead cost and the cost of inspection of a single item, it will be possible to estimate the total cost of inspection. In most cases, the cost of inspection and the limit of error may be suitably balanced so as to keep both within reasonable limits.

5.3 Two-Stage Sampling — The computation of cost in the case of two-stage sampling is somewhat complicated, since in addition to the constant overhead cost, two other types of costs are involved, namely, the cost of sampling the first-stage units and the cost of sampling and testing the items. The formula for calculating the number of items from each first-stage unit (k) so as to get the maximum benefit at minimum cost is as follows:

$$k = \frac{s_w}{s_b}\sqrt{\frac{C_p}{C_s}}$$

where

s_w = standard deviation for the items within the same first-stage unit,

s_b = standard deviation for the first-stage units in the lot,

C_p = cost of sampling a single first-stage unit, and

C_s = cost of sampling and testing a single item.

Thus, more items will have to be taken from a first-stage unit if the variability within a first-stage unit is relatively greater than that between the first-stage units and also if the relative cost of sampling a first-stage unit is greater than that of sampling and testing an item. The number of first-stage units (n) that should be selected can then be determined from the formula in **3.5**. The combination of the values of n and k so obtained will be such that the cost for the given limit of error would be minimum.

6. SELECTION OF SAMPLE

6.1 The sample size obtained as in **3** and **4** would give an estimate of the average quality within the specified limit of error only if the units in the sample are selected at random. To ensure the randomness of selection, IS : 4905-1968* may be used.

APPENDIX A

(*Clause* 2.9)

METHOD OF CALCULATION OF STANDARD DEVIATION

A-1. UNGROUPED DATA

A-1.1 If $x_1, x_2, x_3, \ldots\ldots x_n$ denote a set of n observed values, then the standard deviation (s) is given by the formula:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 \ldots\ldots + (x_n - \bar{x})^2}{n}}$$

$$\text{where} \quad \bar{x} = \frac{x_1 + x_2 + \ldots\ldots\ldots + x_n}{n}$$

$$\text{or} \quad s = \sqrt{\frac{x_1^2 + x_2^2 \ldots\ldots x_n^2 - \dfrac{(x_1 + x_2 + \ldots\ldots + x_n)^2}{n}}{n}}$$

NOTE — In case the number of test results (n) is small, the standard deviation is obtained by dividing the sum of squares of the deviations of the individual test results from their mean value by '$n - 1$' so as to obtain an unbiased estimate of the lot standard deviation.

Example 8:

It is required to find the standard deviation of the following ten observed values of ends per centimetre of a certain cotton fabric:

53·4, 54·5, 52·8, 53·7, 55·2, 54·6, 53·8, 52·9, 54·0, 53·0

For convenience of calculation, the deviations of the actual values from certain fixed value are taken and their standard deviation is calculated which will be the same as in case of the actual values. Taking deviations of the above values from 50, we get:

3·4, 4·5, 2·8, 3·7, 5·2, 4·6, 3·8, 2·9, 4·0, 3·0

*Methods for random sampling.

Therefore, $x_1 + x_2 + \ldots\ldots + x_n = 3{\cdot}4 + 4{\cdot}5 + \ldots\ldots + 3{\cdot}0$
$= 37{\cdot}9$

Squaring the deviations of the values from 50, we get:
11·56, 20·25, 7·84, 13·69, 27·04, 21·16, 14·44, 8·41, 16·00, 9·00

Substituting in the formula, we get:

$$s = \sqrt{\frac{11{\cdot}56 + 20{\cdot}25 + \ldots\ldots + 9{\cdot}000 - \frac{(37{\cdot}9)}{10}}{9}}$$

$$= \sqrt{\frac{149{\cdot}390 - 143{\cdot}641}{9}}$$

$$= \sqrt{\frac{5{\cdot}749}{9}}$$

$$= 0{\cdot}799$$

Hence, the standard deviation of the observed values of ends per centimetre is 0·799 ends per centimetre.

A-2. GROUPED DATA

A-2.1 When the number of observations is large, the formula given in **A-1.1** is too cumbersome for calculation. The data are, then, grouped into a suitable number of groups and the frequency for each group is calculated. If $d_1, d_2, \ldots\ldots d_k$ denote the mid-points of the k groups in which the data are grouped and $f_1, f_2, \ldots\ldots f_k$ represent the frequency in the k groups, then the standard deviation (s) is given by the formula:

$$s = \sqrt{\frac{f_1d_1^2 + f_2d_2^2 + \ldots\ldots + f_kd_k^2}{n} - \frac{(f_1d_1 + f_2d_2 + \ldots\ldots + f_kd_k)^2}{n^2}}$$

$$= \sqrt{\frac{F_2}{n} - \frac{F_1^2}{n^2}}$$

where

$F_2 = f_1d_1^2 + f_2d_2^2 + \ldots\ldots + f_kd_k^2$,
$F_1 = f_1d_1 + f_2d_2 + \ldots\ldots + f_kd_k$, and
$n = f_1 + f_2 + \ldots\ldots + f_k$.

Example 9:

It is required to find the standard deviation and the coefficient of variation of a set of 200 observations of fibre length of wool in

millimetres which are grouped as under:

Group Limit	*Mid-Point of Group (d)*	*Frequency (f)*	*d*	*fd'*	*fd'²*
36 — 40	38	2	−5	−10	50
41 — 45	43	10	−4	−40	160
46 — 50	48	17	−3	−51	153
51 — 55	53	12	−2	−24	48
56 — 60	58	28	−1	−28	28
61 — 65	63 (Z)	35	0	0	0
66 — 70	68	22	+1	+22	22
71 — 75	73	16	+2	+32	64
76 — 80	78	18	+3	+54	162
81 — 85	83	17	+4	+68	272
86 — 90	88	5	+5	+25	125
91 — 95	93	8	+6	+48	288
96 — 100	98	6	+7	+42	294
101 — 105	103	3	+8	+24	192
106 — 110	108	1	+9	+ 9	81
				−153	1 939
				+324	

$n = 200 \qquad F'_1 = 171 \quad F'_2 = 1\,939$

NOTE 1 — The observed values have been grouped as 36 to 40 mm, 41 to 45 mm, 46 to 50 mm, etc. Since each measurement has actually been made to the nearest millimetre, this method of grouping implies that the group ' 36 to 40 mm ' includes all values between 35·5 and 40·5 mm, the group ' 41 to 45 mm ' includes all values between 40·5 to 45·5 mm and so on. The group interval in this case is, therefore, 40·5 — 35·5 = 5·0.

NOTE 2 — For convenience of calculation, it is customary to take deviations of the mid-point from the point with maximum frequency and to take the group interval as the unit. The standard deviation so obtained is then multiplied by the group interval to obtain the standard deviation for the original values.

Standard deviation (s) is given by:

$$s = C \times \sqrt{\frac{F'_2}{n} - \frac{F'^2_1}{n^2}}$$

where

C = Group interval,

$F'_2 = f_1 d'^2_1 + f_2 d'^2_2 + \ldots\ldots + f_k d'^2_k$,

$F'_1 = f_1 d'_1 + f_2 d'_2 + \ldots\ldots + f_k d'_k$, and

$n = f_1 + f_2 + f_3 + \ldots\ldots + f_k$.

In the present case,

$$n = 200,\ F'_1 = 171,\ F'_2 = 1939, \text{ and } c = 5$$

$$\text{Mean} = Z + \frac{F'_1 C}{n}$$

$$= 63 + \frac{171 \times 5}{200}$$

$$= 63 + 4{\cdot}28$$

$$= 67{\cdot}28 \text{ mm}$$

$$s = 5 \times \sqrt{\frac{1939}{200} - \frac{171^2}{200^2}}$$

$$= 5 \times \sqrt{9{\cdot}695 - 0{\cdot}731}$$

$$= 5 \times \sqrt{8{\cdot}964}$$

$$= 5 \times 2{\cdot}994$$

$$= 14{\cdot}97 \text{ mm}$$

$$\text{Coefficient of variation} = 100 \times \frac{\text{standard deviation}}{\text{mean}}$$

$$= 100 \times \frac{14{\cdot}97}{67{\cdot}28}$$

$$= 22{\cdot}25 \text{ percent}$$

BUREAU OF INDIAN STANDARDS

Headquarters:

Manak Bhavan, 9 Bahadur Shah Zafar Marg, NEW DELHI 110002

Telephones: 331 01 31, 331 13 75 Telegrams: Manaksanstha (Common to all Offices)

Regional Offices:		*Telephone*
Central	Manak Bhavan, 9 Bahadur Shah Zafar Marg, NEW DELHI 110002	331 01 31 331 13 75
*Eastern	1/14 C. I. T. Scheme VII M, V. I. P. Road, Maniktola, CALCUTTA 700054	36 24 99
Northern	SCO 445-446, Sector 35-C, CHANDIGARH 160036	2 18 43 3 16 41
Southern	C. I. T. Campus, MADRAS 600113	41 24 42 41 25 19 41 29 16
†Western	Manakalaya, E9 MIDC, Marol, Andheri (East), BOMBAY 400093	6 32 92 95

Branch Offices:

'Pushpak', Nurmohamed Shaikh Marg, Khanpur, AHMADABAD 380001	2 63 48 2 63 49
‡Peenya Industrial Area 1st Stage, Bangalore Tumkur Road BANGALORE 560058	38 49 55 38 49 56
Gangotri Complex, 5th Floor, Bhadbhada Road, T. T. Nagar, BHOPAL 462003	6 67 16
Plot No. 82/83, Lewis Road, BHUBANESHWAR 751002	5 36 27
53/5, Ward No. 29, R.G. Barua Road, 5th Byelane, GUWAHATI 781003	3 31 77
5-8-56C L. N. Gupta Marg (Nampally Station Road), HYDERABAD 500001	23 10 83
R14 Yudhister Marg, C Scheme, JAIPUR 302005	6 34 71 6 98 32
117/418 B Sarvodaya Nagar, KANPUR 208005	21 68 76 21 82 92
Patliputra Industrial Estate, PATNA 800013	6 23 05
T.C. No. 14/1421, University P.O., Palayam TRIVANDRUM 695035	6 21 04 6 21 17

Inspection Offices (With Sale Point):

Pushpanjali, First Floor, 205-A West High Court Road, Shankar Nagar Square, NAGPUR 440010	2 51 71
Institution of Engineers (India) Building, 1332 Shivaji Nagar, PUNE 411005	5 24 35

*Sales Office in Calcutta is at 5 Chowringhee Approach, P. O. Princep Street, Calcutta 700072 27 68 00

†Sales Office in Bombay is at Novelty Chambers, Grant Road, Bombay 400007 89 65 28

‡Sales Office in Bangalore is at Unity Building, Narasimharaja Square, Bangalore 560002 22 36 71

Reprography Unit, BIS, New Delhi, India